

CLAIMS:

1 1. A method for comparing a left formatted file to a right formatted file, comprising:
2 detecting groups of characters in the left and right files;
3 comparing a group in the right file to a corresponding group in the left file to identify a
4 modified group wherein some portion of the group is different between the left file and the right
5 file; and

6 generating a comparison result file containing the modified groups as sections of the
7 comparison result file to maintain the formatting of the modified groups when placed in the
8 comparison result file.

9 2. The method of Claim 1, wherein detecting the groups in the files further
10 comprises detecting and distinguishing tags in the files to determine the groups in the files.

11 3. The method of Claim 2, wherein the files are HTML files and the tags are HTML
12 tags.

13 4. The method of Claim 1 further comprising displaying the comparison result file to
14 a user so that the user views the changed portions of the right file with the formatting intact.

15 5. The method of Claim 1, wherein detecting the groups further comprises
16 normalizing the right file and left file based on one or more rules in a rules database to permit
17 line-by-line comparison of the right and left file despite the formatting in the files.

18 6. The method of Claim 5, wherein the normalization further comprises one or more
19 rules for handling special elements in the files that would inhibit the line-by-line comparison of
20 the file.

21 7. The method of Claim 6, wherein handling the special elements further comprises
22 one or more of removing header tags from the files, removing script references from the files,
23 removing intradocument links from the files, and converting relative URLs into absolute URLs
24 in the file.

25 8. The method of Claim 5, wherein the comparison further comprises comparing the
26 right file to the left file on a line-by-line basis wherein block level HTML elements in each file
27 are treated as separate lines during the comparison.

28 9. The method of Claim 8, wherein the normalization further comprises processing
29 each character of the right and left files.

30 10. The method of Claim 9, wherein the character processing further comprises
31 detecting a preformatting start tag when scanning the document and skipping the pre-formatted
32 text contained between the start tag and a preformatting end tag.

33 11. The method of Claim 10, wherein the character processing further comprises one
34 or more of removing carriage returns, converting multiple white spaces into a single white space,
35 separating block level HTML elements into separate lines by inserting carriage returns before a
36 block level start tag, and keeping text level tags on same line.

12. A system for comparing a left formatted file to a right formatted file, comprising:
means for detecting groups of characters in the left and right files;
means for comparing a group in the right file to a corresponding group in the left file to
identify a modified group wherein some portion of the group is different between the left file and
the right file; and
means for generating a comparison result file containing the modified group as sections
of the comparison result file to maintain the formatting of the modified groups when placed in
the comparison result file.

13. The system of Claim 12, wherein the detecting means further comprises means for
detecting and distinguishing tags in the files to determine the groups in the files.

14. The system of Claim 13, wherein the files are HTML files and the tags are HTML
tags.

15. The system of Claim 12 further comprising means for displaying the comparison
result file to a user so that the user views the changed portions of the right file with the
formatting intact.

16. The system of Claim 12, wherein the detecting means further comprises a
normalizer for processing the right file and left file based on one or more rules in a rules database
to permit line-by-line comparison of the right and left file despite the formatting in the files.

55 17. The system of Claim 16, wherein the normalizer further comprises one or more
56 rules for handling special elements in the files that would inhibit the line-by-line comparison of
57 the file.

58 18. The system of Claim 17, wherein the normalizer further comprises one or more of
59 removing header tags from the files, removing script references from the files, removing
60 intradocument links from the files, and converting relative URLs into absolute URLs in the file.

61 19. The system of Claim 16, wherein the comparing means further comprises means
62 for comparing the right file to the left file on a line-by-line basis wherein each block in each file
63 is treated as a line during the comparison.

64 20. The system of Claim 16, wherein the normalizer further comprises a character
65 processor that processes each character of the right and left files.

66 21. The system of Claim 20, wherein the character processor further comprises means
67 for detecting a preformatting start tag when scanning the document and means for skipping the
68 pre-formatted text contained between the start tag and a preformatting end tag.

69 22. The system of Claim 21, wherein the character processor further comprises one or
70 more of means for removing carriage returns, means for converting multiple white spaces into a
71 single white space, means for separating block level HTML elements into separate lines by
72 inserting carriage returns before a start tag, and means for keeping text level tags on same line.